DOCUMENT RESUME

ED 175 931 TM 009 670

AUTHOR Faddis, Bonnie J.; And Others
TITLE An Empirical Comparison of ESEA Title I Evaluation Models A and B.
PUB DATE Apr 79
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Comparative Analysis; Compensatory Education Programs; Control Groups; *Evaluation Methods; Grade 9; *Models; National Norms; Norm Referenced Tests; Post Testing; Pretesting; *Program Evaluation; Reading Achievement; Remedial Instruction; Remedial Reading; Research Design; Senior High Schools; *Test Results; Writing
IDENTIFIERS Comprehensive Tests of Basic Skills; *Elementary Secondary Education Act Title I; Local Norms; Out of Level Testing; *RMC Models

ABSTRACT

Ninth grade Elementary Secondary Education Act (ESEA) Title I evaluation results in reading and English-writing projects were compared, using the RMC Model A with local empirical norms, Model A with national interpolated norms, and Model B. The Comprehensive Tests of Basic Skills (CTBS)--Total Reading and Total Language subtests--were administered to assess program impact. Students were pretested out-of-level and posttested at-level. Different estimates of program impact resulted, with Model B yielding the most favorable results and Model A local norms yielding the least favorable results. (These differences are discussed in relation to the comparability of models, the appropriateness of each model for the school district tested, and the methods, the appropriateness of each model for the school district tested, and the methods that could be employed to determine the most valid measure of impact.) (Author/MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED175931

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

An Empirical Comparison of ESEA Title I Evaluation Models A and B

Bonnie J. Faddis Judith A. Arter

Northwest Regional Educational Laboratory

Anne Zwertchek

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Bonnie J. Faddis

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM009 670

Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April 8-12, 1979

Abstract

The present study compares ninth grade ESEA Title I evaluation results in reading and English-writing projects using Model A with local empirical norms, Model A with national interpolated norms, and Model B. Different estimates of program impact resulted, with Model B yielding the most favorable results and Model A-Local Norms yielding the least favorable results. These differences are discussed in relation to the comparability of models, the appropriateness of each model for the school district tested, and the methods that could be employed to determine the most valid measure of impact.

Introduction

The Title I Evaluation and Reporting System (TIERS) outlined by Tallmadge and Wood (1976) consists of three basic evaluation models. In Model A, the norm-referenced model, Title I students are compared to students in the norming group who scored at the same percentile on the pretest. If Title I students did not receive the Title I treatment, it is assumed that they would maintain the same percentile on the posttest. Any increase in percentile rank from pretest to posttest is assumed to be due to the Title I program. Model A evaluations can be implemented using either national or locally-developed norms. As Wood and Tallmadge (1976) point out, national norms are generally based on a large sample of students selected to be representative of the nation's school population. These norms would provide a valid no-treatment expectation for a nationally representative sample of Title I students, but would not necessarily be appropriate for local populations. Local groups may differ in one or more educationally relevant ways from the national norm group, and in these cases local norms may provide a more sound no-treatment expectation.

In Model B, the control group model, Title I students are compared to a locally selected group of comparable students. If Title I students did not receive the Title I treatment, it is assumed that they would achieve the same percentile on the posttest as the control group. Any difference between the Title I group and the control group on the posttest is assumed to be due to the Title I program.

3/79
3743A

In Model C, Title I students are compared to a group of higher scoring students. Pupils are assigned to the Title I or comparison group on the basis of a strict cutoff score on the pretest. If students did not receive the Title I treatment, it is assumed that they would have achieved the same percentile on the posttest as a score which is linearly projected downward from the comparison group. Any difference between this projection and the Title I group is assumed to be due to the Title I program.

The User's Guide for the TIERS (Tallmadge & Wood, 1976) asserts that if the models are implemented properly, results across models will be comparable. Thus, data across programs using different models can be aggregated in order to determine a nationwide Title I effect.

Comparability is typically interpreted to mean that if a program was to implement all three models, the measures of impact would be the same. Previous studies investigating whether models are comparable have been of two types. First are simulations in which assumptions are made about the nature of populations and distributions of scores, and simulated data are generated by computer and are then analyzed using procedures prescribed for the various models (Echternacht, 1978; Mandeville, 1978; Stennan & Raffeld, 1977). Second are examinations of actual data obtained from school districts in which two or more models have been implemented simultaneously (Kearns, 1978; Davidoff, 1978; Stennan & Raffeld, 1977). In general, all of the simulations and studies of real data have demonstrated that the models yield different estimates of program impact.

There are, however, limitations to these studies. Although simulations are useful in determining which model is theoretically the best, under certain conditions, and whether the models are in theory equivalent, events may occur in actual practice which are not accounted for by the assumptions made in simulations. Likewise, studies of real data have not demonstrated that all of the models being compared are appropriate for the population and/or have been properly implemented.

In the present study, the impact of a ninth grade Title I program in a large metropolitan school district was assessed using the Model B control group design. Theoretically, this model has the fewest threats to internal validity, and should provide the most precise estimates of program impact when properly implemented. Since data for constructing local norms were available, program impact was also calculated using the Model A with local norms procedures. Finally, although the local testing dates did not correspond to the empirical norming dates, the publisher's interpolated national norms were used to assess project impact with Model A. Issues of comparability of the three models are assessed in light of both the absolute differences in results and the appropriateness of each model for the school district tested.

Method

Selection

Students from four high schools in a large southwestern school district were eligible for ESEA Title I assistance due to the high percentage of students from low income families in those attendance areas. Two high schools which were not eligible for Title I funds, but which still had a large percentage of low income students were selected as control schools. Within each control school, students were identified as "Title I students" using the same criteria for selection as used in the actual Title I schools. With district-developed criterion-referenced tests in the areas of English/writing and reading, remedial-track students were separated into two groups: (1) students with the lowest scores were removed from district classes to receive supplemental "direct" Title I assistance; (2) remedial students who remained in the district-supported classes were considered to be "indirect" Title I students since they received the indirect benefit of reduced class size. Separate evaluations were conducted to assess achievement gains made in classes for direct and for indirect students.

Testing

The Comprehensive Tests of Basic Skills (CTBS) Total Reading and Total Language subtests were used to assess program impact in the reading and English/writing projects respectively. Students were pretested out-of-level with Level 3 in the Fall, 1977, and posttested in-level with Level 4 in the Fall, 1978. Testing was completed during the first two weeks of October for both years. Procedures outlined by Tallmadge and Wood (1976) were used in calculating estimates of program impact with Model B, Model A with local norms, and Model A with national interpolated norms.

Results

Since a large number of the ninth-grade students served by the Title I program dropped out or transferred between the pre and posttests, it is important to consider whether or not the students with complete evaluation data are representative of or comparable to the total group served. Table 1 shows the means and standard deviations of pretest expanded standard scores for direct and indirect Title I and control group students for both subject areas. The most obvious trend is the considerable loss of students from pre to posttest. In the Title I schools, students with both pre and posttest tended to score higher on the pretest than students who were available only for the pretest. One could reasonably assume that the lower-scoring students dropped out at a higher rate than the higher-scoring students. This notion is also supported by the more restricted variance among students with complete data. In the control schools, the trend is less clear. First, a larger proportion of the total control group have complete data than in the Title I group. Instead of less variability among students with complete data, the control students with complete data were more spread out than those with pretest scores only. Mean expanded standard scores increased slightly for control students with complete data, except for direct reading students.

Table 1

Comparability of Pretest Standard Scores for Students with Complete and Incomplete Data

	Title I			Control		
	Mean	SD	n	Mean	SD	n
DIRECT						
CTBS Total Reading						
Pretest Data Only	446.7	57.4	520	446.7	60.4	98
Pre- and Posttest	454.2	54.1	236	443.7	67.2	59
CTBS Total Language						
Pretest Data Only	439.8	65.5	360	424.9	60.2	75
Pre- and Posttest	445.4	63.3	164	433.3	65.6	43
INDIRECT						
CTBS Total Reading						
Pretest Data Only	515.6	79.1	1174	536.2	75.9	212
Pre- and Posttest	531.1	72.9	540	539.9	79.1	150
CTBS Total Language						
Pretest Data Only	505.8	75.8	996	477.2	58.8	113
Pre- and Posttest	515.5	68.5	409	477.5	62.6	75

In order to meet the assumptions underlying Model B implementation, Title I students and Control Group students should be comparable on the pretest. Table 2 indicates that the mean pretest NCE scores differ by more than one NCE in all cases so it is necessary to make a statistical adjustment. Since assignment to Title I and Control groups was not random or random-in-effect, the covariance adjustment was discarded in favor of the principal axis adjustment. All groups except the indirect students in language are within about two NCEs on the pretest.

Comparison of Models

Table 3 shows the pre- and posttest means and standard deviations of expanded scale scores, and the national and local percentiles for direct and indirect Title I and comparison students. Project impact using Model A with national norms, Model A with local norms, and Model B was calculated using the RMC instructions (Tallmadge and Wood, 1976). Results from these models for direct, indirect, and combined Title I students are shows in Table 4.

Although project impact differs across the three models, some fairly consistent patterns emerge. NCE losses are apparent in all models, with Model A-Local Norms showing the greatest losses, followed by Model A-National Norms and Model B. In reading, there is about the same amount of loss for both direct and indirect students as measured by all three models. In language, all three models show greater losses for direct than for indirect students.

Table 2

Comparability of Pretest Scores for Title I and Comparison Group Students

Group		Pretest		NCE Difference T - C
		Percentile	NCE	
CTBS Reading				
Direct	Title I	18.2	30.9	2.1
	Comparison	15.7	28.8	
Indirect	Title I	42.0	45.8	-1.6
	Comparison	45.0	47.4	
D + I	Title I	33.6	41.1	-1.1
	Comparison	35.6	42.2	
CTBS Language				
Direct	Title I	18.4	31.0	1.9
	Comparison	16.1	29.1	
Indirect	Title I	42.5	46.0	8.1
	Comparison	28.3	37.9	
D + I	Title I	35.5	42.2	7.7
	Comparison	23.1	34.5	

Table 3

Pre- and Posttest Scores for Title I and Comparison Students

Students		N	Expanded Standard Scores				National Percentiles		Local Percentiles	
			Pretest		Posttest					
			X	SD	X	SD	Pre	Post	Pre	Post
CTBS Total Reading:										
Direct	Title I	236	454.25	54.13	474.01	56.91	18.2	16.0	9.4	5.4
	Control	59	443.70	67.25	466.95	65.99	15.7	14.2	8.0	4.7
Indirect	Title I	540	531.13	72.90	551.80	79.99	42.0	38.2	30.8	20.8
	Control	150	539.86	79.07	566.13	93.97	45.0	42.7	33.7	25.7
D + I	Title I	776	507.75	76.40	528.14	81.94	33.6	30.0	22.4	14.6
	Control	209	512.71	87.31	538.13	97.68	35.6	33.4	24.8	17.2
CTBS Total Language:										
Direct	Title I	164	445.38	63.31	454.99	69.51	18.4	16.6	9.9	5.4
	Control	43	433.33	65.58	458.21	75.96	16.1	17.1	7.9	5.4
Indirect	Title I	409	515.51	68.54	537.38	76.30	42.5	41.5	31.8	27.6
	Control	75	477.55	62.55	496.19	70.99	28.3	27.7	18.2	14.0
D + I	Title I	573	495.44	74.16	515.80	83.18	35.3	32.9	24.6	18.8
	Control	118	461.43	66.90	483.35	74.80	23.1	24.1	13.6	10.8

Table 4

Comparison of Evaluation Results with Three Models

Analysis	Impact (NCEs)		
	Model A National Norms	Model A Local Norms	Model B
CTBS Total Reading:			
Direct	-1.8	-5.4	-0.5
Indirect	-2.1	-5.6	-0.4
D + I	-2.1	-5.4	-0.9
CTBS Total Language:			
Direct	-1.4	-4.8	-3.5
Indirect	-0.5	-1.9	-0.2
D + I	-1.5	-3.3	-1.6

Since the results from the three evaluation models are not the same, it remains to be shown which model or models are "correct" or most appropriate. The following analyses were undertaken to examine appropriateness:

1. Comparison of Title I and Control group students on pretest scores (Table 2). Comparability would lend support to the appropriateness of Model B.

2. Application of the norm-referenced models to the control group scores (Table 5). Since the control group did not participate in Title I, measured impact should be zero (the no-treatment expectation). If the NCE gain is not zero, then the equal percentile assumption in Model A may not be appropriate in this instance.

3. Verification that the CTBS norming sample includes students who are similar to those in this study. If not, use of Model A with national norms may be inappropriate.

Appropriateness of Models

Results of the above analyes were as follows:

1. As previously noted in Table 2, direct and indirect reading students and direct language students in the Title I and control groups were within about two NCEs of each other on the pretest. This difference is small enough to consider Model B valid when the principal axis method is used to adjust for the pretest differences. However, for indirect language students, the pretest difference for Title I and control group students is considerably larger (8.1 NCEs). Part of this difference may be attributed to the differential loss of students from pre to posttest in the Title I group which caused their pretest mean to be higher (Table 1). However, there still seems to be some lack of comparability between Title I and control group indirect language students.

2. Table 5 shows the NCE gains that would result if Model A were applied to the control group students. Theoretically, these gains should be zero NCEs if the equal percentile assumption of Model A is valid. However, using either national norms or local norms with the control group results in negative gains. Differences between the Title I group NCE losses and the control group NCE losses roughly approximate the results obtained with Model B. That is, although both groups lose ground with respect to the norm groups, the Title I students lose slightly more than the control students, which results in a small negative impact with Model B.

3. According to the Technical Bulletin for CTBS (CTB/McGraw-Hill, 1974), students from minority backgrounds, from large urban public school districts, and from the southwestern United States were included in the norming sample. Thus, the national norms probably are representative of students in this study. However, the national norms do not exclude Title I students, and thus some growth at the lower percentiles is built into the norms. In the local norms, Title I students were excluded, making the local norms more "difficult" for the low-achieving students. However, this does not explain the greater losses found with local norms since Title I students were omitted from both the beginning-of-ninth-grade and the beginning-of-tenth-grade norms.

As a result of these analyses, more confidence is placed in the estimate of impact as measured by Model B. Although the Title I and comparison students are not precisely the same, Model B with a principal axis adjustment for pretest differences probably yields the most accurate results. Fortunately, these are also the "best" results. The equal percentile assumption of Model A is somewhat suspect in these data since the control group shows losses similar to the Title I group with both national and local norms.

Table 5

Comparison of Evaluation Results for Control Group

Analysis	Impact (NCEs)	
	Model A National Norms	Model A Local Norms
CTBS Total Reading:		
Direct	-1.4	-5.2
Indirect	-1.3	-4.3
D + I	-1.2	-4.9
CTBS Total Language:		
Direct	0.9	-2.0
Indirect	-0.4	-3.2
D + I	0.7	-2.2

Conclusions

This paper attempted to examine the claim that results from the TIERS evaluation models are comparable. Results may be considered comparable between programs using different models to the extent that a model is appropriate and correctly implemented in the various programs. Results of comparing Model A with local norms, Model A with national norms, and Model B in the ninth grade Title I program in a large metropolitan high school district showed that results were not the same. It was also shown that the models were not equally appropriate. Finally, it was demonstrated that evaluation results should not be accepted without checking the validity of the design used.

REFERENCES

CTB/McGraw-Hill. Comprehensive tests of basic skills: Technical bulletin No. 1. Monterey, California, 1974.

Davidoff, S. Will aggregation cause aggravation? Paper presented at AERA, Toronto, 1978.

Echternacht, G. The use of different models in the ESEA Title I evaluation system. Paper presented at AERA, Toronto, 1978.

Kearns, D. ESEA Title I evaluation model equivalency and comparability with real data. Paper presented at AERA, Toronto, 1978.

Mandeville, G. K. An evaluation of Title I Model C1 in the special regression model. Paper presented at AERA, Toronto, 1978.

Stennan, D., & Raffeld, P. Is Model A2 equivalent to Model A1? Technical report, Title I Technical Assistance Center, Region 3, Austin, Texas, 1977.

Tallmadge, G. K., & Wood, C. T. User's guide: ESEA Title I evaluation and reporting system. RMC Research Corporation, Mountain View, California, 1976.

Wood, C. T., & Tallmadge, G. K. Local norms: ESEA Title I evaluation and reporting system, Technical Paper No. 7. RMC Research Corporation, Mountain View, California, 1976.